to 1≤i≤K, and K is any integer according to 2≤K≤N, while m is made to be one or more integers according to i≤m*K+i≤N, to extract respective m*K+i-th frames of the first speech.

[0019] With K and i being maintained, the second speech may be generated by extracting the respective m*K+i-th frames of the first speech, as m is incremented between 0 and K−1.

[0020] The score calculator may use acoustic scores of frames of the second speech, calculated by the acoustic model, as determined acoustic scores of respective frames of the first speech that correspond to the frames of the second speech, and derive an acoustic score of one of the frames other than the select frames, as an adjacent frame and being adjacent to one or more of the respective frames of the first speech, based on one or more acoustic scores of the frames of the second speech and/or one or more of determined acoustic scores of the respective frames of the first speech.

[0021] Based on a determined temporal distance between the adjacent frame and two frames of first speech, of the extracted select frames, which are temporally on both sides of the adjacent frame, the score calculator may use, as the acoustic score of the adjacent frame, a determined acoustic score of either one of the two frames as the acoustic score of the adjacent frame or a calculated acoustic score of either one of two corresponding frames of the second speech.

[0022] The score calculator may use, as the acoustic score of the adjacent frame, a statistical value based on determined acoustic scores of two frames of the first speech, of the extracted select frames, which are temporally on both sides of the adjacent frame, or based on calculated acoustic scores of two frames of the second speech corresponding to the two frames of the first speech, or the score calculator may use, as the acoustic score of the adjacent frame, a statistical value obtained by applying a weighted value to each determined acoustic score of the two frames of the first speech, or to each determined acoustic score of the two frames of the second speech, based on respectively determined temporal distances between the adjacent frame and the two frames of the first speech.

[0023] The acoustic model may be trained by using one or more second training speeches respectively generated based on frame sets differently extracted from a same first training speech.

[0024] The preprocessor may be configured to extract the frame sets from the first training speech, generate the one or more second training speeches by respectively using the extracted frame sets, and train the acoustic model by using the generated one or more second training speeches.

[0025] In one general aspect a speech recognition method includes receiving input of first speech to be recognized, extracting some frames from all frames of the first speech, generating a second speech by using the extracted frames, calculating an acoustic score of the second speech by using a Deep Neural Network (DNN)-based acoustic model, and calculating an acoustic score of the first speech based on the calculated acoustic score of the second speech.

[0026] The acoustic model may be a Bidirectional Recurrent Deep Neural Network (BRDNN) acoustic model.

[0027] The extracting of some frames may include extracting select frames from all frames of the first speech according to a predetermined uniform interval, dividing all of the frames of the first speech into two or more groupings and extracting one or more select frames from each of the groupings, or extracting select frames according to an interval that is based on determined signal strengths of frames of the first speech.

[0028] The calculating of the acoustic score of the first speech may include using two acoustic scores of frames of the second speech as acoustic scores of two frames of the first speech that correspond to the two frames of the second speech and using at least one acoustic score of the frames of the second speech for an acoustic score of an adjacent frame, of the first speech, that is adjacent to the two frames of the first speech.

[0029] The calculating of the acoustic score of the first speech may include using an acoustic score of either one of the two frames of the first speech or one of the two frames of the second speech as the acoustic score of the adjacent frame based on a determined temporal distance between the adjacent frame and the two frames of the first speech which are temporally on both sides of the adjacent frame.

[0030] The calculating of the acoustic score of the first speech may include using, as the acoustic score of the adjacent frame, a statistical value of the acoustic scores of the two frames of the first speech or acoustic scores of the two frames of the second speech, or using a statistical value obtained by applying a weighted value to the acoustic scores of the two frames of the first speech, or to the acoustic scores of the two frames of the second speech, based on a determined temporal distance between the adjacent frame and the two frames of the first speech.

[0031] In one general aspect, a speech recognition apparatus includes a frame set extractor configured to extract one or more frame sets, each differently including less than all frames of an input first training speech, a training data generator configured to generate one or more second training speeches by respectively using the extracted one or more frame sets, and a model trainer configured to train the acoustic model by using the generated one or more second training speeches.

[0032] The acoustic model may be a Bidirectional Recurrent Deep Neural Network (BRDNN).

[0033] The apparatus may further include a processor that includes the frame set extractor, the training data generator, and the model trainer, the processor further configured to extract select frames from a first speech of a user for recognition, generate a second speech using the extracted select frames, and recognize the first speech based on calculated acoustic scores of frames, of the first speech and other than the select frames, using acoustic scores of the second speech calculated by the acoustic model.

[0034] The frame set extractor may extract each of an i-th frame set according to m*K+i and from an N number of all of the frames of the first training speech, wherein i is any integer of 1≤i≤K is any integer of 2≤K≤N, and m is any integer of i≤mK+i≤N.

[0035] In one general aspect, a speech recognition method includes extracting one or more frame sets, each differently including less than all frames of an input first training speech, generating one or more second training speeches by using the extracted one or more frame sets, and training the acoustic model by using the generated one or more second training speeches.

[0036] The acoustic model may be a Bidirectional Recurrent Deep Neural Network (BRDNN).

[0037] The extracting of the frame sets may include setting a value i for a reference frame i to be 1, and a value of